Naval Submarine Medical Research Laboratory

NSMRL Report 1125 17 October 1988

BESTSRCH: An automated system for identification of GC/MS peaks

by

D. V. Tappan
E. Heyder
L. W. Mooney

Released by:

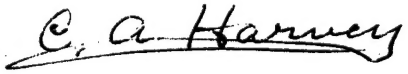
C. A. HARVEY, CAPT, MC, USN
Commanding Officer
Naval Submarine Medical Research Laboratory

BESTSRCH: AN AUTOMATED SYSTEM FOR IDENTIFICATION OF GC/MS PEAKS

D. V. TAPPAN, E. HEYDER AND L. W. MOONEY

NAVAL SUBMARINE MEDICAL RESEARCH LABORATORY
REPORT NO. 1125

APPROVED AND RELEASED BY



C. A. Harvey, CAPT MC USN
Commanding Officer
NavSubMedRschLab

SUMMARY PAGE

THE PROBLEM

To develop software which will satisfactorily remove chromatographic contamination from the apex spectrum of each peak of a Gas Chromatography/Mass Spectrometry (GC/MS) total ion chromatogram and will perform spectral library searches for each peak while the computer is unattended.

FINDINGS

A routine has evolved and has been extensively tested that determines when the best average spectrum for removing background contamination and interference from overlapping chromatographic peaks has been found. Each peak of a total ion chromatogram is corrected for interference, searched against a library of known mass spectra, and the search results are printed without operator intervention.

APPLICATIONS

Either preliminary or optimized searches for the identity of a series of unknown organic chemicals representing all or part of a total ion chromatogram produced during GC/MS analyses may be made. The removal of interferences for a series of chromatographic peaks without help from the analyst allows a major component of data analysis to be performed unattended during off hours when the GC/MS instrumentation and the computer control system are not otherwise in use. The procedure, which has broad general utility, has been applied to the analysis of volatile organic components found in samples of submarine air.

ADMINISTRATIVE INFORMATION

This investigation was conducted under the reimbursable funding from "(U) The Absorption of Volatile Organic Chemicals by Crew Members Located at Different Watch Stations Aboard Submarines." The manuscript was submitted for review on 9 Sept 1988, approved for publication on 17 October 1988, and designated NSMRL Laboratory Report No. 1125.

ABSTRACT

In order to carry out time consuming data reduction processes while simultaneously conducting gas chromatographic/mass spectrometric (GC/MS) analyses with a single computerized control system, it has been necessary to develop software capable of making sophisticated decisions without assistance from the analyst for use during off hours when the computer is not otherwise occupied.

A software routine has been devised and tested that determines the quality of a series of spectral library searches to find the best average spectrum for removing background contamination and the interference of overlapping chromatographic peaks from each peak of a GC/MS total ion chromatogram (TIC). Peak spectra are corrected for interference, searched against a library of known mass spectra, and the results are printed without operator intervention by the software developed.

Either preliminary or optimized searches for the identity of the organic materials separated in an entire chromatogram or in a selected portion of a TIC obtained by GC/MS analysis may be made. The procedure, which has wide general utility, has been applied to the analysis of volatile organic components found in samples of submarine air.

INTRODUCTION

Undoubtedly the most complicated task in the operation of a Gas Chromatography/Mass Spectrometry (GC/MS) laboratory is the identification of compounds from the mass spectra obtained from analyses. Such analyses are especially involved for samples of submarine air studied in this work, which represent the products formed by men and equipment in a closed environment and contain particularly complex mixtures of volatile organic components. Since the air purification equipment of the submarine removes much of the contaminating material, at any moment the variety of organics in the air represents a balance between the continuous air purification and the work/living processes.

While more efficient air purification equipment might be engineered, the more effectively such machinery is designed to remove contaminants, the more costly, complex, and space consuming it becomes. It is therefore very important that detailed and accurate information about air quality be obtained for evaluating the purification process as well as for answering questions about the seriousness or cause of unusual air quality situations that may arise from time to time (1,2). The present work complements the analyses of the major components of air, notably O_2, CO, and CO_2 that are regularly measured as part of normal operating procedures (3).

EXPERIMENTAL METHODS

In these studies, gas chromatographic/mass spectrometric analyses were performed on a Hewlett-Packard (HP) 59970 MS Chem Station using a Model 5970 Mass Selective Detector (MSD) and a HP-5880 Gas Chromatograph. Submarine air samples to be analyzed for volatile organic contaminants were collected by drawing air through glass columns containing glass beads, Tenax-GC, Ambersorb XE-340, and charcoal (Cat. #7142, Envirochem, Inc., Kemblesville, PA) at a rate of about 0.3 liters per minute. The mixed resin collection tubes were sealed in glass containers with teflon lined caps before and after collection. The tubes were stored at -7°C until the samples were analyzed.

Components adsorbed on the collection tubes were eluted by heating the tubes at 275°C in a stream of helium in an Envirochem Unicon Series 310A desorbing and concentrating apparatus which injects the concentrated samples directly onto the gas chromatographic column for separation. The chromatograms discussed were obtained using RTX-5 capillary columns, 0.32 mm ID x 50 meters with 1 micron adsorbent coating (Cat. #10254; Restek Corp., Port Matila, PA). The column temperature was held at 30°C for 8 minutes after sample injection and was then increased by 3 degrees per minute until 150 degrees was reached. The temperature was then increased by 8 degrees per minute to 270 degrees where it was held for 30 minutes to thoroughly clean the column for subsequent analyses. Library searches of MSD spectra were performed using the Hewlett- Packard NBS_REVF spectral library. This report describes the software routines that we have developed for selecting and applying proper cleanup procedures to experimental spectra for analysis.

ANALYTICAL RESULTS

To illustrate the complexity encountered during analysis and identification of the organic components in submarine air, Figure 1 shows a total ion chromatogram (TIC) obtained by GC/MS analysis of 5 liters of air from the Auxiliary Machine Room (AMR) of a 688 class submarine during submergence on a routine training mission. While each air sample obtained from each area of each submarine is unique in composition, the TIC shown in Figure 1 is quite typical with respect to the number and size of chromatographic peaks observed during analyses of air samples obtained from submarines or many indoor environments. In this case, 118 peaks were detected by the integration software (4). Figure 2 shows an expanded 10-minute segment of the same data to better demonstrate the integration procedure used for defining and measuring the area of the individual peaks. Each peak is plotted from the total ion counts, i.e., the sum of the instrumental counts for the individual ions for the 10-30 or more separate spectra which are detected as the analyte enters the mass selective detector from the gas chromatograph. Since the spectra are collected at regular time intervals during the analysis, more spectra are recorded for wider than for narrower peaks.

Figure 3 further illustrates the steps involved in the production of a TIC and the kind of data collected by the computer from the mass selective detector. The TIC shown in the left portion of the figure is a short, expanded section of the chromatogram of Figure 2. The spectra on the right correspond to the data collected by the MSD and saved by the computer at the times marked 1, 2, and 3 on the TIC. It is apparent that the sums of the ion abundances, represented by the vertical spectral lines result in a much higher value on the TIC for the spectrum produced at time 2 than that at time 1. Similarly, the total abundance of mass ions for the spectrum at time 3 is intermediate in size between those of times 1 and 2.

The data presented in Figure 3 represent only 3 spectra of about 1500 spectra that were collected and stored during the analysis of this air sample. Fourteen spectra were stored during the analysis of the peak which has its apex at 35.82 minutes with only 3 of these shown here. Since a value must be stored for each of the points represented by a vertical bar on each spectrum, very large data files are produced which require extensive data reduction and large amounts of storage for archiving.

Figure 4 presents the spectrum obtained at the apex of the peak at 31.01 minutes; this is the first major peak on the left side of Figure 2. Figure 5 shows a printout of the search results, i.e., the identification data and their calculated reliabilities for the spectrum of Figure 4. The method of data presentation used is that of the NBS_REVF reference library package (5) which may be consulted for interpretation of the various pieces of information. The overall reliability values shown in the left data column in the lower section of the printout indicate that the spectrum of octane from the library best matches that at the apex of the peak. The spread between the reliabilities of the first and second match selections strongly suggest the correctness of the first match choice.

The BESTSRCH routine described in this report was developed for automating the tedious process of choosing and correcting the many spectra that are required for the complete analysis of the organic components separated from complex mixtures such as those from air collected from closed environments. The total ion chromatographs and individual spectra of the figures should be referenced to assist in a study of the software routines described below.

DISCUSSION

As may be implied from the data of Figures 1 and 2, the individual organic components from the air samples analyzed in this work can not be completely separated without excessively slow elutions which are very costly in operating time and overburden the data storage capacity of the analytical system. To improve the quality of identification of peaks for data such as those illustrated, correction must be made for overlapping of peaks as well as for background materials arising from the chromatographic column and other parts of the analytical system. Since contamination from one peak into adjacent peaks can not be distinguished from background contamination, the best adjustment that can be made toward obtaining the spectra of pure compounds for identification purposes consists of the subtraction of appropriate spectra or average spectra, on a point by point basis, to remove the contribution of background or contaminating material.

In the following description of the automated search routine that has been developed, it will be noted that considerable effort is made to obtain the least contaminated spectrum for each peak with several searches performed for each peak of the TIC and with several different background corrections used. Because there may be hundreds of peaks in a chromatogram obtained for a single sample, we have tried to develop routines for selecting the best corrected, i.e. background subtracted, spectra without the necessity of operator intervention. Numerous variations on the peak and background selection procedures have been studied with the version shown here the most satisfactory combination of selections devised to date. The routines presented have been under continuous development for at least a year and it is quite possible that further changes may be made as additional experience is accumulated. Despite the possibility of further evolution of the routines, the system is sufficiently useful in its present state to merit reporting and may be applied whenever multiple peak identification is required.

We do not wish to indicate that the automated routines are able to make reasonable selections from current libraries to match all chromatogrphis peaks that may be encountered, especially when the samples analyzed are very complex. Considerable judgment on the part of the analyst is still required for evaluating the final results. For example, it is not uncommon for compounds selected as best matches to be outside the expected ranges of volatility or reactivity for a specific region of a chromatogram. We agree with the manufacturer of the analytical system that no routine search procedure can completely displace the judgment of a trained anlayst (5). However, having pointed out the potential problems of the method, we hasten to indicate that 80-90% of the spectra the chromatograms searched have been satisfactorily identified by the system.

The automated search procedure will first be briefly summarized and then will be described step-by-step with each phase of the procedure annotated in the software printout included. The language of the software is the MACRO language of the HP-59970 Chem Station (4).

Summary of the Search Program, BESTSRCH. The TIC is first integrated (4) and then each of the integrated peaks that meets a specified minimum size requirement is subjected to a series of searches to locate the best match(es) for the corrected experimental spectrum against the spectra stored in the reference library (5). Background corrections are made by subtracting the average spectrum from the valley between the peaks on the wider side of the peak then on the narrow side of the peak and finally from both sides of the peak. If the peak is more than 10 spectra wide on either side, the 10th spectrum from the apex is used as the outside limit of the peak for purposes of background subtraction. The background windows are set to include 1 to 3 spectra on either side of the apex, including the outside or limit spectrum, depending on the number of spectra included in the peak.

The first search is made for an average spectrum obtained for the three spectra which include the apex spectrum and one spectrum on each side of the apex. The match quality index for the best match of a library spectrum with the experimental spectrum is determined and retained for comparison against the results of later searches. In order to expedite the searching of the total set of peaks, the searches of any particular peak are stopped if a match quality of 90 or above is obtained. Such matches are considered essentially perfect by the developers of the library software (5).

The second search is made for the average of 1 to 3 spectra, including the apex spectrum, on the wider side of the peak. The background spectrum to be subtracted is an average of the spectra in the background area on the wider side. The third search is analogous to the second but is made for the spectra on the narrow side of the peak. The fourth search is made for the spectra surrounding the apex with an average background subtracted for the window areas on both sides of the peak. All library match qualities and the peak and background areas of the cumulative best search are retained.

If the selected match quality (90) with a library spectrum has not been attained after four searches, a more rigorous search is performed using the peak and background regions from the best of the four exploratory searches. The strategy for this search includes tilting of experimental spectrum to compensate for differences in response between the mass spectrometers used to collect the library spectra and the experimental spectra. The results of the fifth search are retained as the most satisfactory available unless, as occasionally happens, an earlier search gave a better result. In such a case, the search strategy is reset, and the best earlier search is repeated.

Detailed Description of BESTSRCH. The following discussion annotates the various sections of the software routine included in the next section which documents the searching procedures. The numbered paragraphs correspond to the numbers shown in the left column of the software printout. To simplify this description, some of the "housekeeping" steps, especially those involving the moving of data between storage registers are not explicitly discussed.

1. N = peak number to be analyzed. NPEAKS = Total number of peaks found by the software integrator in the setup routines, AUTOSRCH or BTCHSRCH (auto or batch search). These preliminary data processing programs, which set operational parameters and both require BESTSRCH to perform the analyses, are described elsewhere (6).

2. The printer is set not to waste paper between printouts.

3. Initialize variables and search strategy parameters. Quick library searches, which save time and often locate satisfactory matches with library spectra, are used initially. NUM_HITS = number of possible library matches found for a peak spectrum. QLY = quality of best library match (0-100). Q1-Q4 = variable locations for saving quality values from various searches. NH1-NH4 = variables for saving NUM_HITS from various searches. LB1-LB4 = variables for saving the library entry number for the best match for the individual searches.

4. Mostly self explanatory definitions of parameters found for each peak. RET_TIME = retention time for spectrum at apex of peak.

5. For a peak to be analyzed, it must be larger than the minimum peak size, MPK, established in setup routines. S1-S3 are scan numbers of the spectra at the start time, retention time, and end time for the peak being analyzed.

6. NS = Number of Scans across entire peak. DS (Delta Scan) = time from one scan to the next.

7. Peak must contain at least 5 scans to be analyzed. SK (Skew) = 0 if more scans follow than precede the maximum (apex) scan for the peak; otherwise SK = 1.

8. SS (Small Side) = S2 - S1, number of scans: peak scan - first scan. LS (Large Side) = S3 - S2, number of scans: last scan - peak scan.

9. Swap SS and LS, if necessary, to match Small Side and Large Side of peak.

10. Providing NS > LS (ie, NS is greater than LS), if SS or LS > 10, the sides of the peaks are reduced to 10 scans for analysis to avoid subtracting backgrounds that are too far from the apex. T1 and T3, the start time and end time of the peak, are adjusted accordingly.

11. If SS > 5 then SBW (Small Background Window) is set to (SS/3)- 1; otherwise it's 0. If SS > 8, SPW (Small Peak Window) = (SS/3)-2. Windows for large side of curve, LBW and LPW, are set similarly. Windows for averaging spectra across peaks or for subtracting background are thus restricted to 0-2 spectra for background and 0-1 for the peak spectra on either side of the apex.

12. A1 and A2 enclose a region at the top of a peak that includes one spectrum on each side of the apex. The average of the spectra from A1 to A2 is calculated and a library search on the result is performed.

13. If the quality of the best match with a library spectrum is greater than or equal to 90, then the searches are stopped (MORE is set to 0). If the quality is < 90 and the peak is not too narrow or asymmetric, searching continues.

14. Set limits of peak to include one spectrum on small side of the peak and LPW spectra on large side, X1 to X2. Background, Z1 to Z2 is set for LBW spectra on large side of peak.

15. Average peak scans and background scans and subtract background. Perform library search.

16. If quality of best library match is >= 90 then set MORE to stop the searches; otherwise if quality of the present search is better than that of the previous search, save peak scan and background scan limits in A1, A2, B1, and B2.

17. Skip next search if small side of peak (SS) contains < 3 spectra.

18. Set limits for analysis of peak, X1 and X2, and for background, Z1 and Z2, on small side of peak analogously to those set for large side of peak for previous search.

19. Obtain average peak and background spectra, subtract background, and perform library search.

20. If quality of match with a library spectrum is better than for previous searches, save the peak and background limits for further analysis. If library match quality is >= 90 then stop the searches (MORE = 0); otherwise continue searching.

21. Get average of the background spectra on both the small and large sides of the peak in the SBW and LBW regions. Add the two average background spectra and multiply by 0.5 to obtain composite average.

22. Calculate the average of the three spectra at the apex. Subtract the combined average background spectrum from average peak spectrum and do a library search on the result.

23. Note whether the match quality of this search is better than that of previous searches. If the match quality is >= 90, set the stop code, MORE.

24. Get either the spectrum from the last search, if its quality was better than earlier searches, or get the average peak and background spectra from the search that provided the best match between peak spectrum and a library spectrum.

25. Set search strategy to do a more comprehensive search of the spectral library to involve tilting the experimental spectrum to compensate for differing spectrometer responses. Perform library search to improve the quality of previous best match.

26. Write an error message if the peak was too narrow or too asymmetric for a satisfactory library search.

27. More error messages for peaks unsuitable for searching.

28. Draw picture of the peak spectrum if it could not be searched or if no match was found.

29. Retrieve spectra from library for the three best matches with the corrected experimental spectrum. Fewer spectra are retrieved if less than three possible matches were found.

30. Print the screen containing the experimental spectrum and spectra of the best library matches along with information about the quality of the searches.

31. Stop the searches if the qualities of the comprehensive search was better than or equal to that of the best preliminary search. If the quality of the comprehensive search was lower than that of the best earlier search, reset the search strategy for a quick search to retrieve the information from the best quick search.

32. Print data for the five best library matches for the best quick search to allow confirmation or rejection of the information from the comprehensive search.

33. Repeat the searches for each integrated peak of the total ion chromatogram.

REFERENCES

1. Tappan, D.V., D.R. Knight, E. Heyder, and P.K. Weathersby. Volatile Organic Components of Air Samples Collected from Vertical Launch Missile Capsules. NAVSUBMEDRSCHLAB Memo Report 1124, 1988.

2. Callahan, A.B., D.V. Tappan, and W.L. Mooney. Analysis of Hydraulic Fluids and Lubricating Oils for the Formation of Trimethylpropane Phospate (TMP-P). Final Report Prepared for SEA 05R23, 1988.

3. Panel on Monitoring, Subcommittee on Air Quality, Committee on Toxicology, Board on Environmental Studies and Toxicology, National Research Council. Monitoring the Air in Submarines. National Academy Press. Washington, D.C. 1988.

4. HP 59970C Chem Station, Automating Your Analyses With Macros. Publication No. 59970-90017. Hewlett-Packard Co., Palo Alto, CA. 1986.

5. HP 59970C Chem station, PBM Search and Parametric Retrieval Software Handbook. Publication No. 59973-90003. Hewlett-Packard Co., Palo Alto, CA. 1986.

6. Tappan, D.V., E. Heyder, and W.L. Mooney. NSMRL's Software Library for Collection and Analysis of Gas Chromatography/Mass Spectrometry Data. In Preparation.

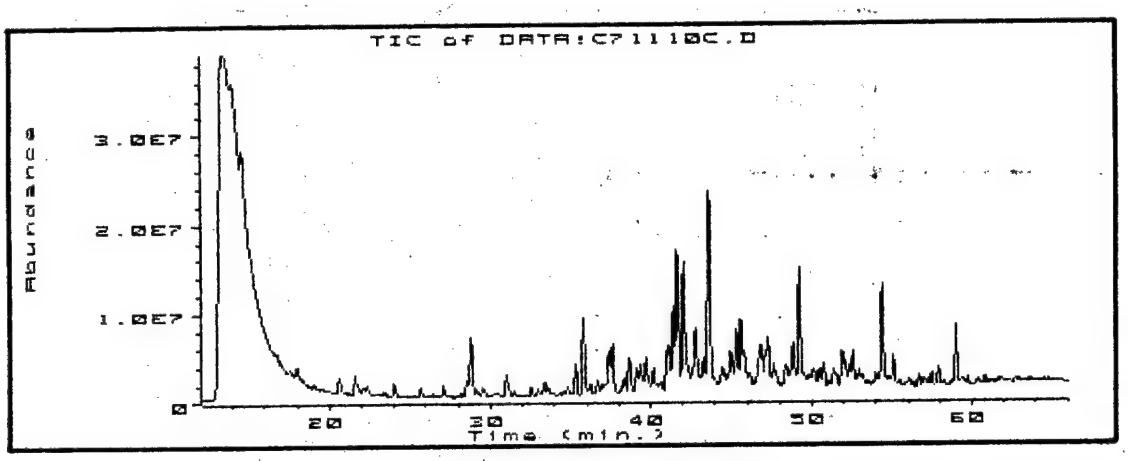


FIGURE 1. Total ion chromatogram obtained from GC/MS analysis of volatile organic components trapped from an air sample collected from an operating submarine during submergence.

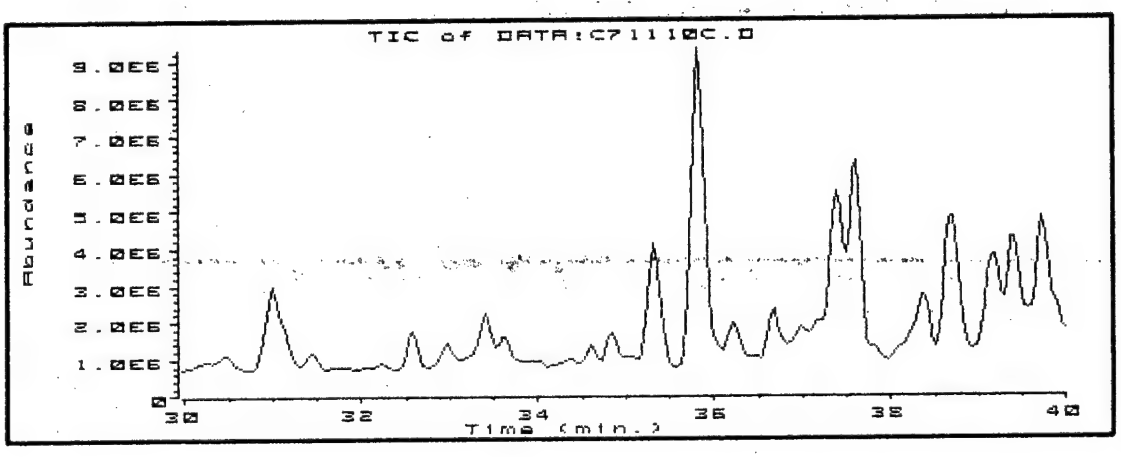


FIGURE 2. An expanded ten minute segment of the total ion chromatogram shown in Fig. 1 to better illustrate the separation of individual peaks. See text for further explanation.

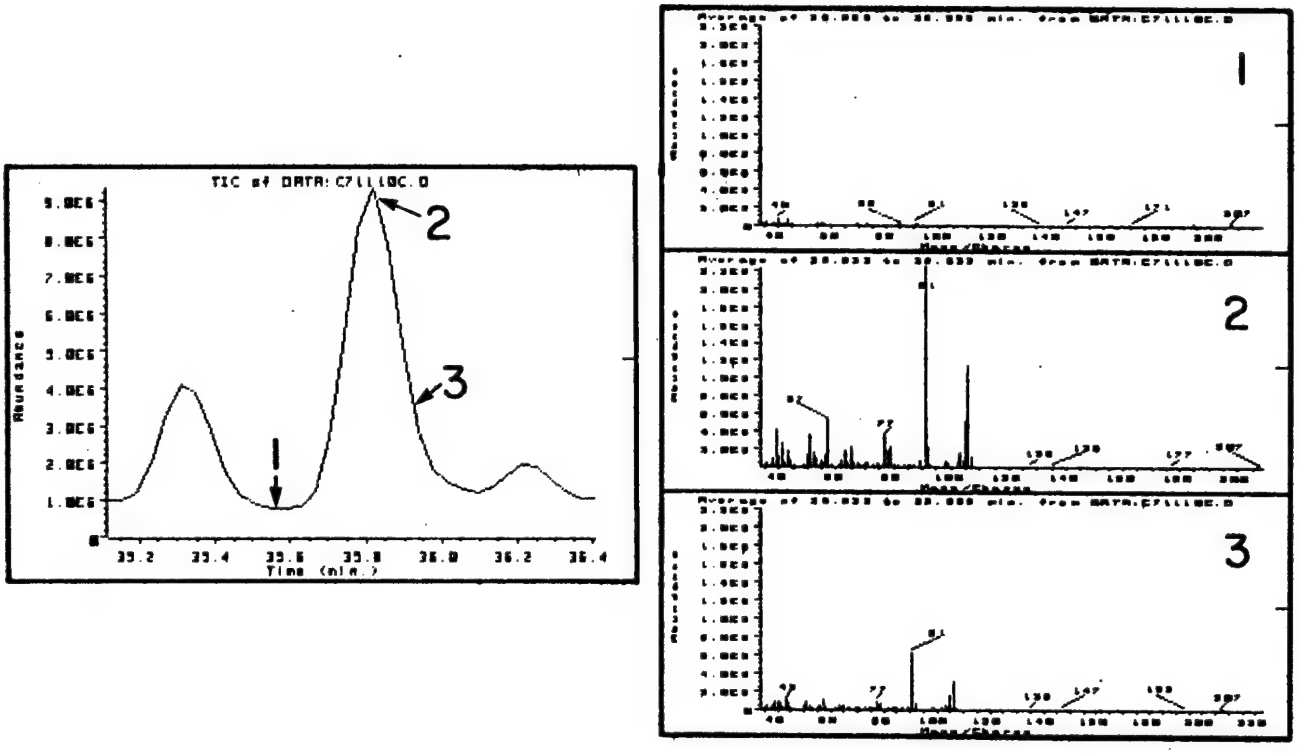


FIGURE 3. A greatly expanded segment of the chromatogram shown in Figures 1 and 2 with the mass spectra collected for points 1, 2, and 3 which are marked on the TIC. See text for further explanation.

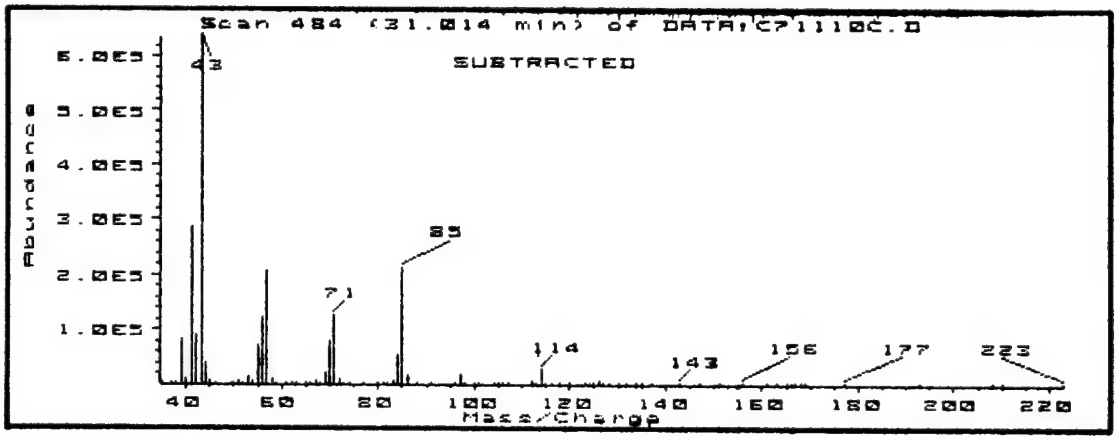


FIGURE 4. Mass/charge spectrum at the apex of peak at 31.01 minutes from Figures 1 annd 2.

Average of 30.976 to 31.052 min. from DATA:C71110C.D
HELENA 20 OCT 87;UNDRWAY-RECIRC,DAY 8;AMR 2045-2100

	Name	MolWt	Formula
1.	Octane (DOT)(8CI9CI)	114	C8H18
2.	Heptane, 2,4-dimethyl- (8CI9CI)	128	C9H20
3.	Hexane, 3-ethyl- (8CI9CI)	114	C8H18
4.	Hexane, 2,4-dimethyl- (8CI9CI)	114	C8H18
5.	Hexane, 2,3,4-trimethyl- (8CI9CI)	128	C9H20
6.	Hexane, 2-methyl- (8CI9CI)	100	C7H16
7.	Heptane, 2,3-dimethyl- (8CI9CI)	128	C9H20
8.	Heptane, 3-methyl- (8CI9CI)	114	C8H18
9.	Hexane, 1-propoxy- (9CI)	144	C9H20O
10.	Cyclobutane, 1,2-diethyl- (9CI)	112	C8H16
11.	Aziridine, 1,2,3-trimethyl-, trans- (8CI	85	C5H11N
12.	Nonane, 2,6-dimethyl- (8CI9CI)	156	C11H24
13.	Pentane, 3-ethyl-2-methyl- (8CI9CI)	114	C8H18
14.	4-Piperidinemethanamine (9CI)	114	C6H14N2
15.	Hexane, 3,3-dimethyl- (8CI9CI)	114	C8H18
16.	1-Hexene, 4,5-dimethyl- (8CI9CI)	112	C8H16
17.	1-Heptene, 6-methyl- (8CI9CI)	112	C8H16
18.	1-Hexene, 3,4,5-trimethyl- (9CI)	126	C9H18
19.	Aziridine, 2,2-dimethyl- (8CI9CI)	71	C4H9N
20.	Hydroxylamine, O-(3-methylbutyl)- (9CI)	103	C5H13NO

	Prob	CAS#	Ref#	K	dK	Flag	Tilt	%	Con	C_I	R_IV
1.	*89	00111-65-9	2552	81	12	1	2	99	8	62	80
2.	59	02213-23-2	4136	64	28	1	0	93	24	27	30
3.	*55	00619-99-8	2568	50	40	1	0	100	35	20	39
4.	*54	00589-43-5	2560	51	47	2	0	70	24	22	26
5.	*32	00921-47-1	4125	26	49	0	0	45	40	10	18
6.	29	00591-76-4	1382	50	36	2	0	100	43	8	17
7.	27	03074-71-3	4141	45	46	2	0	100	37	10	13
8.	*26	00589-81-1	2562	23	53	1	0	49	45	8	14
9.	25	53685-78-2	6655	35	48	2	0	100	48	7	12
10.	*25	61141-83-1	2285	21	45	2	0	21	48	7	13
11.	*25	00693-88-9	574	35	65	3	0	103	48	7	13
12.	*20	17302-28-2	8733	28	48	1	0	27	54	5	15
13.	*20	00609-26-7	2567	33	58	1	0	79	51	5	18
14.	*20	07144-05-0	2459	26	58	3	0	21	51	5	13
15.	20	00563-16-6	2555	47	43	2	0	78	51	5	15
16.	*18	16106-59-5	2277	32	40	0	0	26	56	4	27
17.	*15	05026-76-6	2262	25	41	1	0	18	59	3	14
18.	*15	56728-10-0	3789	23	47	1	0	28	59	3	14
19.	*15	02658-24-4	237	27	52	2	0	22	59	3	14
20.	15	19411-65-5	1568	45	36	2	0	71	59	3	15

FIGURE 5. Identification data produced by the program, BESTSRCH, for the chromatographic peak from Figures 1 and 2 at 31.01 minutes.

Printout of software.

```
NAME BESTSRCH
   N = 0
   STRAT 1,6,NO,SCAN
1) WHILE N<NPEAKS
2)  WRITELN 701,"*27&12P"
3)  N=N+1
    NUM_HITS = 0
    QLY = 0
    Q1 = -1
    Q2 = -1
    Q3 = -1
    Q4 = -1
    Q5 = -1
    NH = 0
    NH1 = 0
    NH2 = 0
    NH3 = 0
    NH4 = 0
    LB1 = 0
    LB2 = 0
    LB3 = 0
    LB4 = 0
    MORE = 1
    WRITELN 701,<------------------------------------------------->
    PEAK N
    ROLL -1
4)  T1=START_TIME
    T2=RET_TIME
    T3=END_TIME
5)  IF PEAK_AREA > MPK THEN
      S2 = SCAN_NUM
      SP T1
      S1 = SCAN_NUM
      ROLL -1
      SP T3
      S3 = SCAN_NUM
      ROLL -1
6)    NS = S3 - S1 + 1
      DS = (T3-T1)/(S3-S1)
7)    IF NS > 5 THEN
        SK= 0
8)      SS = S2-S1
        LS = S3-S2
10)     IF SS  > LS THEN
          SK= 1
          SS = LS
          LS = S2 - S1
        ENDIF
11)     IF NS > LS THEN
          IF SS > 10 THEN
            SS = 10
```

```
            IF SK = 0 THEN
              T1 = T2 - (DS*SS)
            ELSE
              T3 = T2 + (DS*SS)
            ENDIF
          ENDIF
          IF LS  > 10 THEN
            LS = 10
            IF SK = 0 THEN
              T1 = T2 - (DS*LS)
            ELSE
              T3 = T2 + (DS*LS)
            ENDIF
          ENDIF
          SBW = 0
          IF SS > 5 THEN
            SBW = (SS/3) - 1
          ENDIF
          SPW = 0
          IF SS > 8 THEN
            SPW = (SS/3) - 2
          ENDIF
          LBW = 0
          IF LS > 5 THEN
            LBW = (LS/3) - 1
          ENDIF
          LPW = 0
          IF LS > 8 THEN
            LPW = (LS/3) - 2
          ENDIF
12)       A1 = T2 - DS
          A2 = T2 + DS
          B1 = 0
          B2 = 0
          AV A1:A2
          CL
          DR
          MES SEARCH 1.,,20,10
          MES PEAK # = ,N,1,19
          MES .. PEAK TIME = ,T2,18,19
          NORM
          PBM NBS_REVF
          GETS RES,X
          Q1 = QUALITY
          NH1 = NUM_HITS
          LB1 = ENTRY_NUM
          ROLL -1
          QLY = QUALITY
          QSCH = 1
13)       IF Q1 >= 90 THEN
            MORE = 0
          ENDIF
          IF MORE > 0 THEN
```

```
            IF LS <  3 THEN
              WRITELN 701,**NARROW OR ASYMMETRIC PEAK**PARTIAL SEARCH MADE**
            ELSE
              IF SK = 0 THEN
14)             X1 = T2 - DS
                X2 = T2 + (DS*LPW)
                Z1 = T3 - (D6*LBW)
                Z2 = T3
              ELSE
                X1 = T2 - (DS*LPW)
                X2 = T2 + DS
                Z1 = T1
                Z2 = T1 + (DS*LBW)
              ENDIF
15)           AV X1:X2
              AV Z1:Z2
              SUB SUPP
              NORM
              CL
              DR
              MES SEARCH 2. PROBABILITY = ,QLY,20,10
              MES PEAK # ,N,1,19
              MES .. PEAK TIME = ,T2,16,19
              PBM NBS_REVF
              GETS RES,X
              Q2 = QUALITY
              NH2 = NUM_HITS
              LB2 = ENTRY_NUM
              ROLL -1
            ENDIF
          ENDIF
          IF Q2 >= QLY THEN
            QLY = Q2
            QSCH = 2
            A1 = X1
            A2 = X2
            B1 = Z1
            B2 = Z2
          ENDIF
16)       IF Q2 >= 90 THEN
            MORE = 0
          ENDIF
          IF MORE > 0 THEN
17)         IF SS < 3 THEN
              WRITELN 701,**NARROW OR ASYMMETRIC PEAK**PARTIAL SEARCH MADE**
            ELSE
              IF SK = 0 THEN
18)             X1 = T2 - (DS*SPW)
                X2 = T2 + DS
                Z1 = T1
                Z2 = T1 + (DS*SBW)
              ELSE
```

```
                X1 = T2 - DS
                X2 = T2 + (DS*SPW)
                Z1 = T3 - (DS*SBW)
                Z2 = T3
              ENDIF
19)           AV X1:X2
              AV Z1:Z2
              SUB SUPP
              NORM
              CL
              DR
              MES SEARCH 3. PROBABILITY = ,QLY,20,10
              MES PEAK # ,N,1,19
              MES .. PEAK TIME = ,T2,16,19
              PBM NBS_REVF
              GETS RES,X
              Q3 = QUALITY
              NH3 = NUM_HITS
              LB3 = ENTRY_NUM
              ROLL -1
            ENDIF
          ENDIF
20)       IF Q3 >= QLY THEN
            QLY = Q3
            QSCH = 3
            A1 = X1
            A2 = X2
            B1 = Z1
            B2 = Z2
          ENDIF
          IF Q3 >= 90 THEN
            MORE = 0
          ENDIF
          IF MORE > 0 THEN
            IF SK = 0 THEN
21)           AV T1:T1+(DS*SBW)
              AV T3-(DS*LBW):T3
            ELSE
              AV T3-(DS*SBW):T3
              AV T1:T1+(DS*LBW)
            ENDIF
            ADD
            MULT 0.5
22)         AV T2-DS:T2+DS
            EX
            SUB SUPP
            NORM
            DR
            MES SEARCH 4. PROBABILITY = ,QLY,20,10
            MES PEAK # = ,N,1,19
            MES .. PEAK TIME = ,T2,18,19
            PBM NBS_REVF
```

```
            GETS RES,X
            Q4 = QUALITY
            NH4 = NUM_HITS
            LB4 = ENTRY_NUM
            ROLL -1
          ENDIF
23)       IF Q4 >= QLY THEN
            QLY = Q4
            QSCH = 4
          ENDIF
          IF Q4 >= 90 THEN
            MORE = 0
          ENDIF
          IF MORE > 0 THEN
24)         IF QSCH = 4 THEN
              ROLL
            ELSE
              AV A1:A2
              IF B1 > 0 THEN
                AV B1:B2
                SUB SUPP
              ENDIF
            ENDIF
            NORM
            DR
            MES SEARCH 5. PROBABILITY = ,QLY,20,10
            MES PEAK # = ,N,1,19
            MES .. PEAK TIME = ,T2,18,19
25)         STRAT 1,6,SMART,SCAN
            PBM NBS_REVF
            GETS RES,X
            Q5 = QUALITY
            ROLL -1
          ENDIF
        ELSE
26)       WRITELN 701,**PEAK TOO ASYMMETRIC TO SEARCH**
          IF PEAK_AREA > MPK THEN
            SP T2
          ENDIF
        ENDIF
      ELSE
        WRITELN 701,**PEAK TOO NARROW**
        IF PEAK_AREA > MPK THEN
          SP T2
        ENDIF
      ENDIF
    ENDIF
    NH = NUM_HITS
    IF NH < 1 THEN
27)   IF PEAK_AREA < MPK THEN
        WRITELN 701,**PEAK TOO SMALL FOR ANALYSIS**
```

```
      ELSE
        WRITELN 701,**NO MATCHES FOUND**
      ENDIF
      SP T2
      CL
      DR
28)   SCR 440,260
    ELSE
      CL
      ROLL
      N1 = 3
29)   IF N1 > NH THEN
        N1 = NH
      ENDIF
      N2 = N1 + 1
      RET N1
      N1 = N1 - 1
      WHILE N1 > 0 DO
        EX
        RET N1
        EX Y,Z
        MER
        N1 = N1 - 1
      ENDWHILE
      EX
      MER
30)   DR 3
      MES Q1: ,Q1,0,0
      MES .NH: ,NH1,7,0
      MES LM: ,LB1,71,0
      MES Q2: ,Q2,0,1
      MES .NH: ,NH2,7,1
      MES LM: ,LB2,71,1
      MES Q3: ,Q3,0,2
      MES .NH: ,NH3,7,2
      MES LM: ,LB3,71,2
      MES Q4: ,Q4,0,3
      MES .NH: ,NH4,7,3
      MES LM: ,LB4,71,3
      MES PK# ,N,0,4
      MES .,,,9,4
      MES TM: ,T2,71,4
      SZ = PEAK_AREA/1E8
      MES SZ: ,SZ,71,5
      GR
      EXT N2
      TAB RES,PRINTER:
      IF MORE = 0 THEN
        CL
        ROLL -1
      ELSE
```

```
31)     IF QLY <= 5 THEN
          CL
          ROLL -1
        ELSE
          EX
          ROLL -1
          CL
          DR
          MES SEARCH 6. PROBABILITY = ,QLY,20,10
          MES PEAK # = ,N,1,19
          MES .. PEAK TIME = ,T2,18,19
          STRAT 1,6,NO,SCAN
          PBM NBS_REVF
          GETS RES,X
          CL
          N1 = 5
          IF NUM_HITS < N1 THEN
            N1 = NUM_HITS
          ENDIF
          N2 = 1
          WHILE N2 <=  N1
            GETS RES,X,,N2
            YY = N2 + 5
32)         MES ** FROM BEST QUICK SEARCH **,,10,2
            MES QUAL       LIB#      CAS#          MOL WT,,4,4
            MES ,QUALITY,3,YY
            MES ,ENTRY_NUM,14,YY
            MES ,CAS_NUM,24,YY
            MES ,MOL_WT,38,YY
            N2 = N2 + 1
          ENDWHILE
          GR
        ENDIF
      ENDIF
    ENDIF
    ROLL -1
    CL
33) ENDWHILE
    WRITELN 701,"#27&166P"
    QUIT
```

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release. Distribution is unlimited
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NSMRL Report 1125	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Naval Submarine Medical Research Laboratory 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Naval Medical Research and Development Command
6c. ADDRESS (City, State, and ZIP Code) Box 900, Naval Submarine Base New Lodnon Groton, CT 06349-5900	7b. ADDRESS (City, State, and ZIP Code) Naval Medical Command, National Capital Region, Bethesda, MD 20814-5044
8a. NAME OF FUNDING / SPONSORING ORGANIZATION Same as 7a. 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 6.1 / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification) BESTSRCH: An Automated System for Identification of GC/MS Peaks

12. PERSONAL AUTHOR(S)
Tappan, D.V., Heyder, E. and Mooney, L. W.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Interim	FROM _____ TO _____	17 October 1988	18

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Gas Chromatography; Mass Spectrometry; Data Management; Submarine Medicine; Volatile Organic Pollutants; Environmental Toxicology

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

In order to carry out very time consuming data reduction processes while simultaneously conducting gas chromatographic/mass spectrometric (GC/MS) analyses with a single computerized control system, it has been necessary to develop software capable of making sophisticated decisions without assistance from the analyst for use during off hours when the computer is not otherwise occupied.

A software routine has been devised and well tested that determines the quality of a series of spectral library searches to find the best average spectrum for removing background contamination and the interference of

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Susan D. Monty, Publications Office	22b. TELEPHONE (Include Area Code) (203) 449-3967 / 22c. OFFICE SYMBOL 421

DD Form 1473, JUN 86 *Previous editions are obsolete.* S/N 0102-LF-014-6603

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

overlapping chromatographic peaks from each peak of a GC/MS total ion chromatogram (TIC). Peak spectra are corrected for interference, searched against a library of known mass spectra, and the results are printed without operator intervention by the software developed.

Either preliminary or optimized searches for the identity of the organic materials separated in an entire chromatogram or in a selected portion of a TIC obtained by GC/MS analysis may be made. The procedure, which has wide general utility, has been applied to the analysis of volatile organic components found in samples of submarine air.